If you say that in a ReLU neural network the gating from the ReLU functions can only split regions, while the linear mapping from the weight matrices can both split and merge regions. Is that reasonable?

Yes—that's a very reasonable and actually quite useful way to think about ReLU networks. Let me sharpen it slightly so it's both precise and insightful.

---

1. ReLU gating: *splitting only*

A ReLU applies an elementwise rule:

$x \longrightarrow \max(0, x)$

Geometrically, each neuron introduces a **hyperplane boundary** (where its pre-activation = 0) that partitions the input space.

- On one side → neuron is **active** (passes signal through)
- On the other → neuron is **inactive** (output = 0)

This means:

- The space is **partitioned into regions** based on activation patterns
- Each region corresponds to a fixed binary gating pattern (a diagonal matrix D of 0s and 1s)

So ReLU:

- **creates new region boundaries**
- **refines (splits) existing regions**
- but does **not map two distinct regions into one**

Why not merge? Because:

- It is **piecewise linear but not mixing across regions**
- Each region keeps its identity via a distinct gating pattern

---

2. Linear maps (weight matrices): *split + merge*

A linear transformation:

$x \longrightarrow Wx$

does something quite different geometrically:

(a) It can split regions

- By stretching/rotating space
- Making previously aligned regions separable in later layers

(b) It can merge regions

This is the key point:

- If two different inputs $x1 \neq x2$ satisfy $Wx1 = Wx2$, then they are **merged**
- This happens whenever W is **not injective** (e.g., rank-deficient, dimensionality reduction)

Even if full-rank:

- It can still **collapse distinctions relevant to future gating**
- After transformation, different regions may fall into the *same side* of future ReLU boundaries

So linear layers:

- **mix information**
- **superimpose regions**
- **can destroy distinctions**

---

3. Combined picture (the important insight)

A ReLU network alternates:

- **ReLU (D):** creates a combinatorial partition of space
- **Linear (W):** reshapes, mixes, and sometimes collapses those regions

So each layer:

$x \longrightarrow DWx$

leads to:

- ReLU → **adds boundaries (splitting)**

- Linear → **warps + possibly merges regions**

---

4. Intuition via “region flow”

You can think of the network as:

- ReLU = *decision maker* → “which region are you in?”
- Linear = *transport/mixer* → “where does this region go next?”

Across layers:

- Regions get **progressively subdivided** (via gating)
- But also **recombined or aligned** (via linear maps)

---

5. Subtle but important refinement

Your statement is *almost* perfectly correct, but one nuance:

> ReLU doesn’t explicitly merge regions, but it can **collapse variation within a region** (e.g., all negative inputs → 0)

So:

- It **kills information** (non-invertible)
- But does not *actively fuse distinct regions into a single structured region* the way linear maps do

---

6. Clean summary

- **ReLU gating:**
  → splits space into more regions
  → defines discrete activation patterns
  → does not mix regions
- **Linear maps:**
  → can split (via geometric distortion)
  → can merge (via projection / mixing)
  → transport regions across space

---

7. Why this viewpoint matters (connects to your earlier ideas)

This directly ties into your earlier themes:

- The **“hash-map / switch statement” view**:
  - ReLU = selects a region (key)
  - Linear = computes within that region
- The **WD (synthetic matrix) view**:
  - Each region has its own effective matrix S = DW
  - Linear layers allow different regions to **share or collide in representation space**
- The **capacity story**:
  - Splitting ReLU → increases number of contexts
  - Merging W → allows reuse and generalization

Final verdict

Your statement is not only reasonable—it’s a **very powerful abstraction**:

> ReLU layers *differentiate* regions, while linear layers *mix and sometimes identify them.*